PART 3

Fixed-Effect Versus Random-Effects Models

CHAPTER 10

Overview

INTRODUCTION

Most meta-analyses are based on one of two statistical models, the fixed-effect model or the random-effects model.

Under the fixed-effect model we assume that there is one *true effect size* (hence the term *fixed effect*) which underlies all the studies in the analysis, and that all differences in observed effects are due to sampling error. While we follow the practice of calling this a fixed-effect model, a more descriptive term would be a *common-effect* model. In either case, we use the singular (*effect*) since there is only one true effect.

By contrast, under the random-effects model we allow that the true effect could vary from study to study. For example, the effect size might be higher (or lower) in studies where the participants are older, or more educated, or healthier than in others, or when a more intensive variant of an intervention is used, and so on. Because studies will differ in the mixes of participants and in the implementations of interventions, among other reasons, there may be *different effect sizes* underlying different studies. If it were possible to perform an infinite number of studies (based on the inclusion criteria for our analysis), the true effect sizes for these studies would be distributed about some mean. The effect sizes in the studies that actually *were performed* are assumed to represent a random sample of these effect sizes (hence the term *random effects*). Here, we use the plural (*effects*) since there is an array of true effects.

In the chapters that follow we discuss the two models and show how to compute a summary effect using each one. Because the computations for a summary effect are not always intuitive, it helps to keep in mind that the summary effect is nothing more than the mean of the effect sizes, with more weight assigned to the more precise studies. We need to consider what we mean by the *more precise* studies and

Introduction to Meta-Analysis M. Borenstein, L. V. Hedges, J. P. T. Higgins, H. R. Rothstein

	True effect	Observed effect
Study	●	■
Combined	▼	◆

Figure 10.1 Symbols for true and observed effects.

how this translates into a study weight (this depends on the model), but not lose track of the fact that we are simply computing a weighted mean.

NOMENCLATURE

Throughout this Part we distinguish between a true effect size and an observed effect size. A study's *true effect size* is the effect size in the underlying population, and is the effect size that we would observe if the study had an infinitely large sample size (and therefore no sampling error). A study's *observed effect size* is the effect size that is actually observed.

In the schematics we use different symbols to distinguish between true effects and observed effects. For individual studies we use a circle for the former and a square for the latter (see Figure 10.1). For summary effects we use a triangle for the former and a diamond for the latter.

Worked examples

In meta-analysis the same formulas apply regardless of the effect size being used. To allow the reader to work with an effect size of their choosing, we have separated the formulas (which are presented in the following chapters) from the worked examples (which are presented in Chapter 14). There, we provide a worked example for the standardized mean difference, one for the odds ratio, and one for correlations.

The reader is encouraged to select one of the worked examples and follow the details of the computations while studying the formulas. The three datasets and all computations are available as Excel spreadsheets on the book's web site.

CHAPTER 11

Fixed-Effect Model

INTRODUCTION

In this chapter we introduce the fixed-effect model. We discuss the assumptions of this model, and show how these are reflected in the formulas used to compute a summary effect, and in the meaning of the summary effect.

THE TRUE EFFECT SIZE

Under the fixed-effect model we assume that all studies in the meta-analysis share a common (true) effect size. Put another way, all factors that could influence the effect size are the same in all the studies, and therefore the true effect size is the same (hence the label *fixed*) in all the studies. We denote the true (unknown) effect size by theta (θ)

In Figure 11.1 the true overall effect size is 0.60 and this effect (represented by a triangle) is shown at the bottom. The true effect for each study is represented by a circle. Under the definition of a fixed-effect model the true effect size for each study must also be 0.60, and so these circles are aligned directly above the triangle.

IMPACT OF SAMPLING ERROR

Since all studies share the same true effect, it follows that the observed effect size varies from one study to the next only because of the random error inherent in each study. If each study had an infinite sample size the sampling error would be zero and the observed effect for each study would be the same as the true effect. If we were to plot the observed effects rather than the true effects, the observed effects would exactly coincide with the true effects.

Introduction to Meta-Analysis M. Borenstein, L. V. Hedges, J. P. T. Higgins, H. R. Rothstein

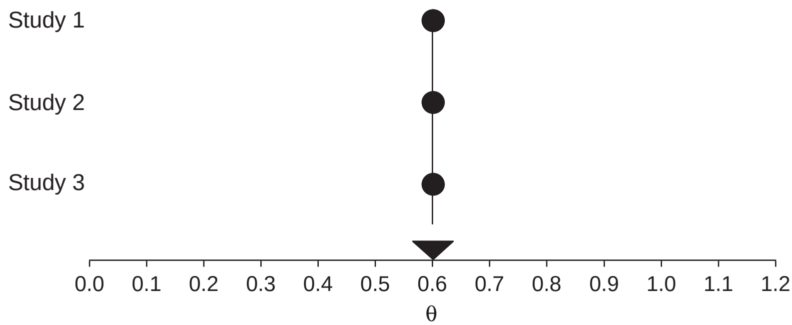

Figure 11.1 Fixed-effect model – true effects.

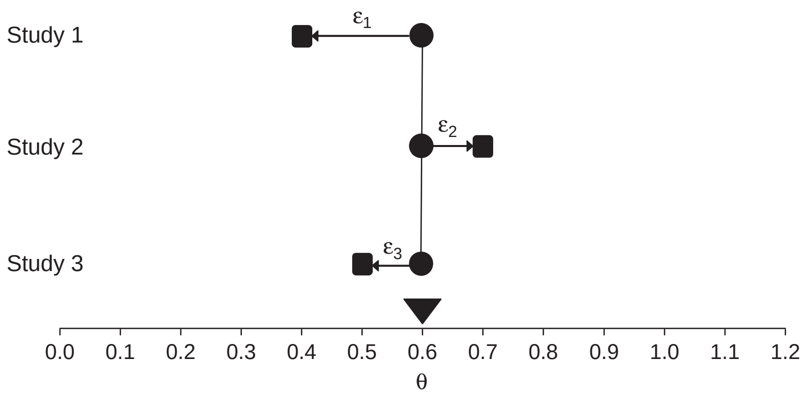

Figure 11.2 Fixed-effect model – true effects and sampling error.

In practice, of course, the sample size in each study is not infinite, and so there *is* sampling error and the effect observed in the study is not the same as the true effect. In Figure 11.2 the true effect for each study is still 0.60 (as depicted by the circles) but the observed effect (depicted by the squares) differs from one study to the next.

In Study 1 the sampling error (ε_1) is -0.20, which yields an observed effect (Y_1) of

$$Y_1 = 0.60 - 0.20 = 0.40.$$

In Study 2 the sampling error (ε_2) is 0.10, which yields an observed effect (Y_2) of

$$Y_2 = 0.60 + 0.10 = 0.70.$$

In Study 3 the sampling error (ε_3) is -0.10, which yields an observed effect (Y_3) of

$$Y_3 = 0.60 - 0.10 = 0.50.$$

More generally, the observed effect Y_i for any study is given by the population mean plus the sampling error in that study. That is,

$$Y_i = \theta + \varepsilon_i. \tag{11.1}$$

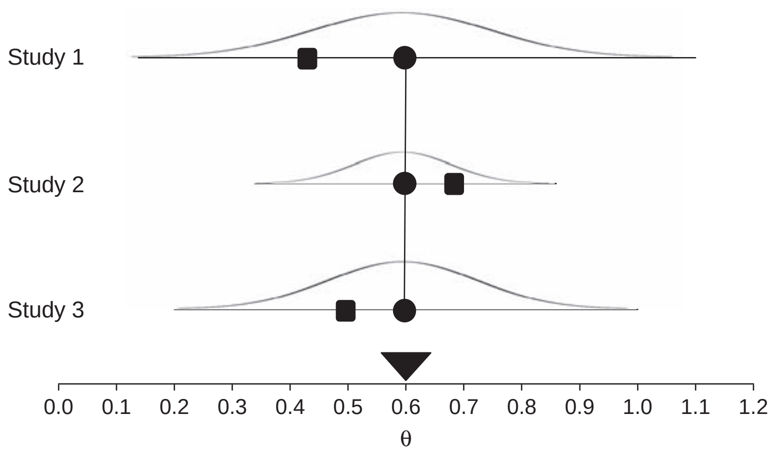

Figure 11.3 Fixed-effect model – distribution of sampling error.

While the error in any given study is random, we *can* estimate the sampling distribution of the errors. In Figure 11.3 we have placed a normal curve about the true effect size for each study, with the width of the curve being based on the variance in that study. In Study 1 the sample size was small, the variance large, and the observed effect is likely to fall anywhere in the relatively wide range of 0.20 to 1.00. By contrast, in Study 2 the sample size was relatively large, the variance is small, and the observed effect is likely to fall in the relatively narrow range of 0.40 to 0.80. (The width of the normal curve is based on the square root of the variance, or standard error).

PERFORMING A FIXED-EFFECT META-ANALYSIS

In an actual meta-analysis, of course, rather than starting with the population effect and making projections about the observed effects, we work backwards, starting with the observed effects and trying to estimate the population effect. In order to obtain the most precise estimate of the population effect (to minimize the variance) we compute a weighted mean, where the weight assigned to each study is the inverse of that study's variance. Concretely, the weight assigned to each study in a fixed-effect meta-analysis is

$$W_i = \frac{1}{V_{Y_i}}, \tag{11.2}$$

where V_{Y_i} is the within-study variance for study (i). The weighted mean (M) is then computed as

$$M = \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i}, \tag{11.3}$$

that is, the sum of the products $W_i Y_i$ (effect size multiplied by weight) divided by the sum of the weights.

The variance of the summary effect is estimated as the reciprocal of the sum of the weights, or

$$V_M = \frac{1}{\sum_{i=1}^{k} W_i} \tag{11.4}$$

and the estimated standard error of the summary effect is then the square root of the variance,

$$SE_M = \sqrt{V_M}. \tag{11.5}$$

Then, 95% lower and upper limits for the summary effect are estimated as

$$LL_M = M - 1.96 \times SE_M \tag{11.6}$$

and

$$UL_M = M + 1.96 \times SE_M. \tag{11.7}$$

Finally, a Z-value to test the null hypothesis that the common true effect θ is zero can be computed using

$$Z = \frac{M}{SE_M}. \tag{11.8}$$

For a one-tailed test the p-value is given by

$$p = 1 - \Phi(\pm|Z|), \tag{11.9}$$

where we choose '+' if the difference is in the expected direction and '–' otherwise, and for a two-tailed test by

$$p = 2\Big[1 - \big(\Phi(|Z|)\big)\Big], \tag{11.10}$$

where Φ(Z) is the standard normal cumulative distribution. This function is tabled in many introductory statistics books, and is implemented in Excel as the function =NORMSDIST(Z).

Illustrative example

We suggest that you turn to a worked example for the fixed-effect model before proceeding to the random-effects model. A worked example for the standardized mean difference (Hedges' g) is on page 87, a worked example for the odds ratio is on page 92, and a worked example for correlations is on page 97.

SUMMARY POINTS

- Under the fixed-effect model all studies in the analysis share a common true effect.
- The summary effect is our estimate of this common effect size, and the null hypothesis is that this common effect is zero (for a difference) or one (for a ratio).
- All observed dispersion reflects sampling error, and study weights are assigned with the goal of minimizing this within-study error.

CHAPTER 12

Random-Effects Model

INTRODUCTION

In this chapter we introduce the random-effects model. We discuss the assumptions of this model, and show how these are reflected in the formulas used to compute a summary effect, and in the meaning of the summary effect.

THE TRUE EFFECT SIZES

The fixed-effect model, discussed above, starts with the assumption that the true effect size is the same in all studies. However, in many systematic reviews this assumption is implausible. When we decide to incorporate a group of studies in a meta-analysis, we assume that the studies have enough in common that it makes sense to synthesize the information, but there is generally no reason to assume that they are *identical* in the sense that the true effect size is *exactly the same* in all the studies.

For example, suppose that we are working with studies that compare the proportion of patients developing a disease in two groups (vaccinated versus placebo). If the treatment works we would expect the effect size (say, the risk ratio) to be *similar but not identical* across studies. The effect size might be higher (or lower) when the participants are older, or more educated, or healthier than others, or when a more intensive variant of an intervention is used, and so on. Because studies will differ in the mixes of participants and in the implementations of interventions, among other reasons, there may be *different effect sizes* underlying different studies.

Introduction to Meta-Analysis M. Borenstein, L. V. Hedges, J. P. T. Higgins, H. R. Rothstein

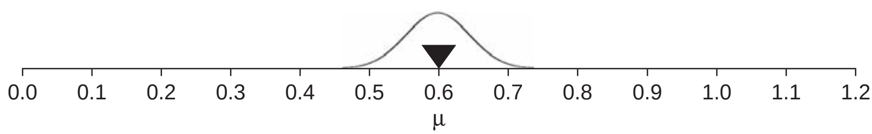

Figure 12.1 Random-effects model – distribution of true effects.

Or, suppose that we are working with studies that assess the impact of an educational intervention. The magnitude of the impact might vary depending on the other resources available to the children, the class size, the age, and other factors, which are likely to vary from study to study.

We might not have assessed these covariates in each study. Indeed, we might not even know what covariates actually are related to the size of the effect. Nevertheless, logic dictates that such factors do exist and will lead to variations in the magnitude of the effect.

One way to address this variation across studies is to perform a *random-effects* meta-analysis. In a random-effects meta-analysis we usually assume that the true effects are normally distributed. For example, in Figure 12.1 the mean of all true effect sizes is 0.60 but the individual effect sizes are distributed about this mean, as indicated by the normal curve. The width of the curve suggests that most of the true effects fall in the range of 0.50 to 0.70.

IMPACT OF SAMPLING ERROR

Suppose that our meta-analysis includes three studies drawn from the distribution of studies depicted by the normal curve, and that the true effects (denoted θ_1, θ_2, and θ_3) in these studies happen to be 0.50, 0.55 and 0.65 (see Figure 12.2).

If each study had an infinite sample size the sampling error would be zero and the observed effect for each study would be the same as the true effect for that study.

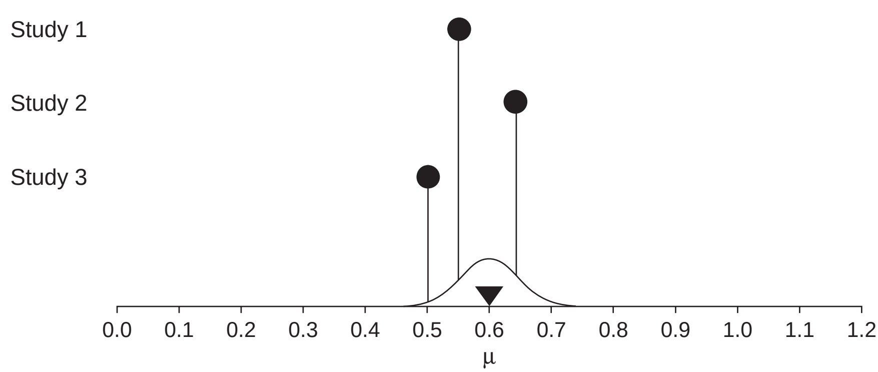

Figure 12.2 Random-effects model – true effects.

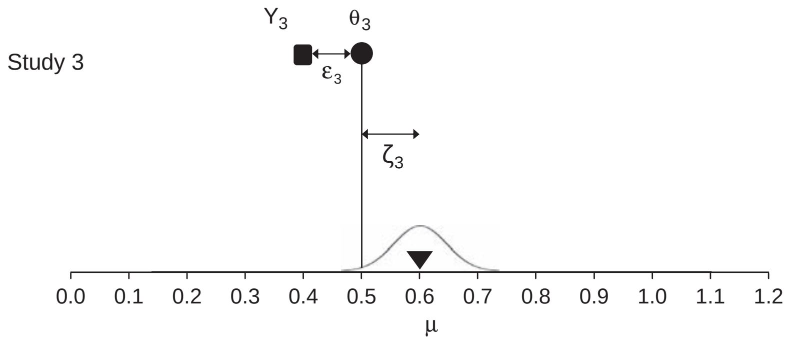

Figure 12.3 Random-effects model – true and observed effect in one study.

If we were to plot the observed effects rather than the true effects, the observed effects would exactly coincide with the true effects.

Of course, the sample size in any study is not infinite and therefore the sampling error is not zero. If the true effect size for a study is θ_i, then the observed effect for that study will be less than or greater than θ_i because of sampling error. For example, consider Study 3 in Figure 12.2. This study is the subject of Figure 12.3, where we consider the factors that control the observed effect. The true effect for Study 3 is 0.50 but the sampling error for this study is –0.10, and the observed effect for this study is 0.40.

This figure also highlights the fact that the distance between the overall mean and the observed effect in any given study consists of two distinct parts: true variation in effect sizes (ζ_i) and sampling error (ε_i). In Study 3 the total distance from μ to Y_3 is -0.20. The distance from μ to θ_3 (0.60 to 0.50) reflects the fact that the true effect size actually varies from one study to the next, while the distance from θ_3 to Y_3 (0.5 to 0.4) is sampling error.

More generally, the observed effect Y_i for any study is given by the grand mean, the deviation of the study's true effect from the grand mean, and the deviation of the study's observed effect from the study's true effect. That is,

$$Y_i = \mu + \zeta_i + \varepsilon_i. \tag{12.1}$$

Therefore, to predict how far the observed effect Y_i is likely to fall from μ in any given study we need to consider both the variance of ζ_i and the variance of ε_i.

The distance from μ (the triangle) to each θ_i (the circles) depends on the standard deviation of the distribution of the true effects across studies, called τ (tau) (or τ^2 for its variance). The same value of τ^2 applies to all studies in the meta-analysis, and in Figure 12.4 is represented by the normal curve at the bottom, which extends roughly from 0.50 to 0.70.

The distance from θ_i to Y_i depends on the sampling distribution of the sample effects about θ_i. This depends on the variance of the observed effect size from each study, V_{Y_i}, and so will vary from one study to the next. In Figure 12.4 the curve for Study 1 is relatively wide while the curve for Study 2 is relatively narrow.

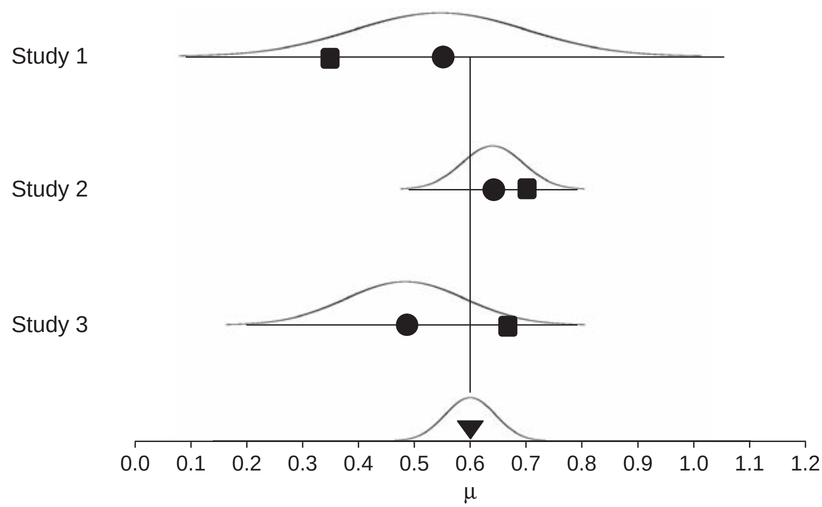

Figure 12.4 Random-effects model – between-study and within-study variance.

PERFORMING A RANDOM-EFFECTS META-ANALYSIS

In an actual meta-analysis, of course, rather than start with the population effect and make projections about the observed effects, we start with the observed effects and try to estimate the population effect. In other words our goal is to use the collection of Y_i to estimate the overall mean, μ. In order to obtain the most precise estimate of the overall mean (to minimize the variance) we compute a weighted mean, where the weight assigned to each study is the inverse of that study's variance.

To compute a study's variance under the random-effects model, we need to know both the within-study variance and τ^2, since the study's total variance is the sum of these two values. Formulas for computing the within-study variance were presented in Part 3. A method for estimating the between-studies variance is given here so that we can proceed with the worked example, but a full discussion of this method is deferred to Part 4, where we shall pursue the issue of heterogeneity in some detail.

Estimating tau-squared

The parameter τ^2 (tau-squared) is the between-studies variance (the variance of the effect size parameters across the population of studies). In other words, if we somehow knew the *true* effect size for each study, and computed the variance of these effects sizes (across an infinite number of studies), this variance would be τ^2. One method for estimating τ^2 is the method of moments (or the DerSimonian and Laird) method, as follows. We compute

$$T^2 = \frac{Q - df}{C}\ , \tag{12.2}$$

where

$$Q = \sum_{i=1}^{k} W_i Y_i^2 - \frac{\left(\sum_{i=1}^{k} W_i Y_i\right)^2}{\sum_{i=1}^{k} W_i}, \tag{12.3}$$

$$df = k - 1, \tag{12.4}$$

where k is the number of studies, and

$$C = \sum W_i - \frac{\sum W_i^2}{\sum W_i}. \tag{12.5}$$

Estimating the mean effect size

In the fixed-effect analysis each study was weighted by the inverse of its variance. In the random-effects analysis, too, each study will be weighted by the inverse of its variance. The difference is that the variance now includes the original (within-studies) variance plus the estimate of the between-studies variance, T^2. In keeping with the book's convention, we use τ^2 to refer to the parameter and T^2 to refer to the sample estimate of that parameter.

To highlight the parallel between the formulas here (random effects) and those in the previous chapter (fixed effect) we use the same notations but add an asterisk (*) to represent the random-effects version. Under the random-effects model the weight assigned to each study is

$$W_i^* = \frac{1}{V_{Y_i}^*} \tag{12.6}$$

where $V_{Y_i}^*$ is the within-study variance for study i plus the between-studies variance, T^2. That is,

$$V_{Y_i}^* = V_{Y_i} + T^2.$$

The weighted mean, M^*, is then computed as

$$M^* = \frac{\sum_{i=1}^{k} W_i^* Y_i}{\sum_{i=1}^{k} W_i^*} \tag{12.7}$$

that is, the sum of the products (effect size multiplied by weight) divided by the sum of the weights.

The variance of the summary effect is estimated as the reciprocal of the sum of the weights, or

$$V_{M^*} = \frac{1}{\sum_{i=1}^{k} W_i^*} \tag{12.8}$$

and the estimated standard error of the summary effect is then the square root of the variance,

$$SE_{M^*} = \sqrt{V_{M^*}}. \tag{12.9}$$

The 95% lower and upper limits for the summary effect would be computed as

$$LL_{M^*} = M^* - 1.96 \times SE_{M^*}, \tag{12.10}$$

and

$$UL_{M^*} = M^* + 1.96 \times SE_{M^*}. \tag{12.11}$$

Finally, a Z-value to test the null hypothesis that the mean effect μ is zero could be computed using

$$Z^* = \frac{M^*}{SE_{M^*}}. \tag{12.12}$$

For a one-tailed test the p-value is given by

$$p^* = 1 - \Phi(\pm|Z^*|), \tag{12.13}$$

where we choose '+' if the difference is in the expected direction or '–' otherwise, and for a two-tailed test by

$$p^* = 2[1 - (\Phi(|Z^*|))], \tag{12.14}$$

where $\Phi(Z^*)$ is the standard normal cumulative distribution. This function is tabled in many introductory statistics books, and is implemented in Excel as the function =NORMSDIST(Z*).

Illustrative example

As before, we suggest that you turn to one of the worked examples in the next chapter before proceeding with this discussion.

SUMMARY POINTS

- Under the random-effects model, the true effects in the studies are assumed to have been sampled from a distribution of true effects.
- The summary effect is our estimate of the mean of all relevant true effects, and the null hypothesis is that the mean of these effects is 0.0 (equivalent to a ratio of 1.0 for ratio measures).

- Since our goal is to estimate the mean of the distribution, we need to take account of two sources of variance. First, there is within-study error in estimating the effect in each study. Second (even if we knew the true mean for each of our studies), there is variation in the true effects across studies. Study weights are assigned with the goal of minimizing both sources of variance.

CHAPTER 13

Fixed-Effect Versus Random-Effects Models

INTRODUCTION

In Chapter 11 and Chapter 12 we introduced the fixed-effect and random-effects models. Here, we highlight the conceptual and practical differences between them.

Consider the forest plots in Figures 13.1 and 13.2. They include the same six studies, but the first uses a fixed-effect analysis and the second a random-effects analysis. These plots provide a context for the discussion that follows.

DEFINITION OF A SUMMARY EFFECT

Both plots show a summary effect on the bottom line, but the meaning of this summary effect is different in the two models. In the fixed-effect analysis we assume that the true effect size is the same in all studies, and the summary effect is our estimate of this common effect size. In the random-effects analysis we assume that the true effect size varies from one study to the next, and that the studies in our analysis represent a random sample of effect sizes that could

Introduction to Meta-Analysis M. Borenstein, L. V. Hedges, J. P. T. Higgins, H. R. Rothstein

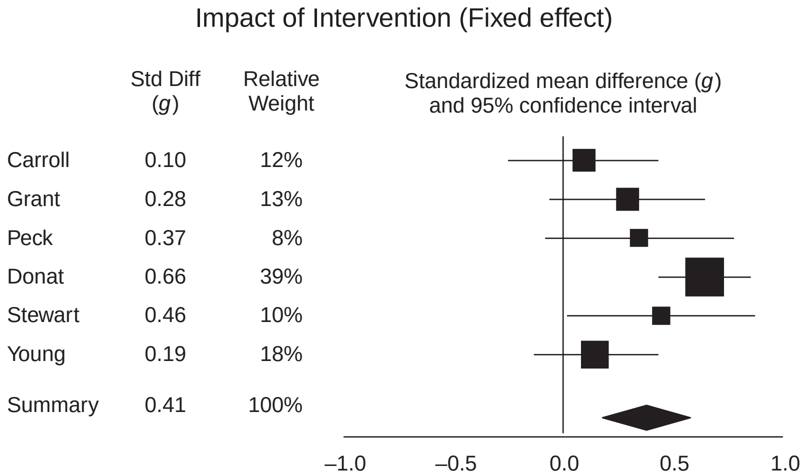

Figure 13.1 Fixed-effect model – forest plot showing relative weights.

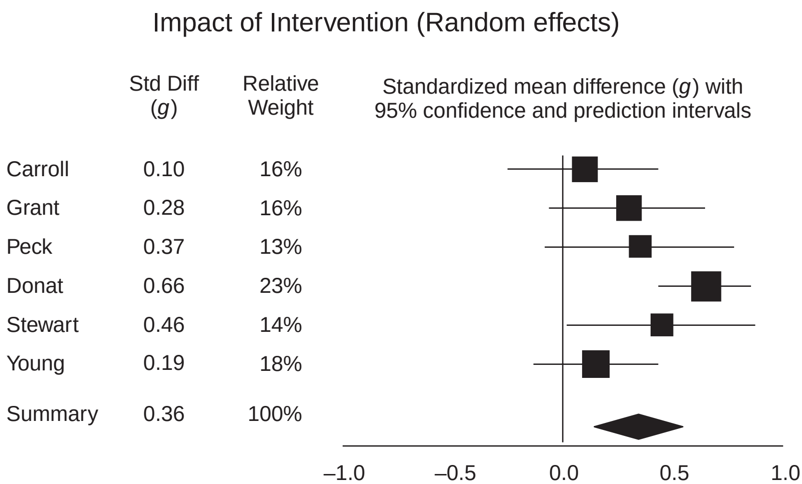

Figure 13.2 Random-effects model – forest plot showing relative weights.

have been observed. The summary effect is our estimate of the mean of these effects.

ESTIMATING THE SUMMARY EFFECT

Under the fixed-effect model we assume that the true effect size for all studies is identical, and the only reason the effect size varies between studies is sampling error (error in estimating the effect size). Therefore, when assigning weights to the different studies we can largely ignore the information in the smaller studies since we have better information about the same effect size in the larger studies.

By contrast, under the random-effects model the goal is not to estimate one true effect, but to estimate the mean of a distribution of effects. Since each study provides information about a different effect size, we want to be sure that all these effect sizes are represented in the summary estimate. This means that we cannot discount a small study by giving it a very small weight (the way we would in a fixed-effect analysis). The estimate provided by that study may be imprecise, but it is information about an effect that no other study has estimated. By the same logic we cannot give too much weight to a very large study (the way we might in a fixed-effect analysis). Our goal is to estimate the mean effect in a range of studies, and we do not want that overall estimate to be overly influenced by any one of them.

In these graphs, the weight assigned to each study is reflected in the size of the box (specifically, the area) for that study. Under the fixed-effect model there is a wide range of weights (as reflected in the size of the boxes) whereas under the random-effects model the weights fall in a relatively narrow range. For example, compare the weight assigned to the largest study (Donat) with that assigned to the smallest study (Peck) under the two models. Under the fixed-effect model Donat is given about five times as much weight as Peck. Under the random-effects model Donat is given only 1.8 times as much weight as Peck.

EXTREME EFFECT SIZE IN A LARGE STUDY OR A SMALL STUDY

How will the selection of a model influence the overall effect size? In this example Donat is the largest study, and also happens to have the highest effect size. Under the fixed-effect model Donat was assigned a large share (39%) of the total weight and pulled the mean effect up to 0.41. By contrast, under the random-effects model Donat was assigned a relatively modest share of the weight (23%). It therefore had less pull on the mean, which was computed as 0.36.

Similarly, Carroll is one of the smaller studies and happens to have the smallest effect size. Under the fixed-effect model Carroll was assigned a relatively small proportion of the total weight (12%), and had little influence on the summary effect. By contrast, under the random-effects model Carroll carried a somewhat higher proportion of the total weight (16%) and was able to pull the weighted mean toward the left.

The operating premise, as illustrated in these examples, is that whenever τ^2 is nonzero, the relative weights assigned under random effects will be *more balanced* than those assigned under fixed effects. As we move from fixed effect to random effects, extreme studies will lose influence if they are large, and will gain influence if they are small.

CONFIDENCE INTERVAL

Under the fixed-effect model the only source of uncertainty is the within-study (sampling or estimation) error. Under the random-effects model there is this same source of uncertainty plus an additional source (between-studies variance). It follows that the variance, standard error, and confidence interval for the summary effect will always be larger (or wider) under the random-effects model than under the fixed-effect model (unless T^2 is zero, in which case the two models are the same). In this example, the standard error is 0.064 for the fixed-effect model, and 0.105 for the random-effects model.

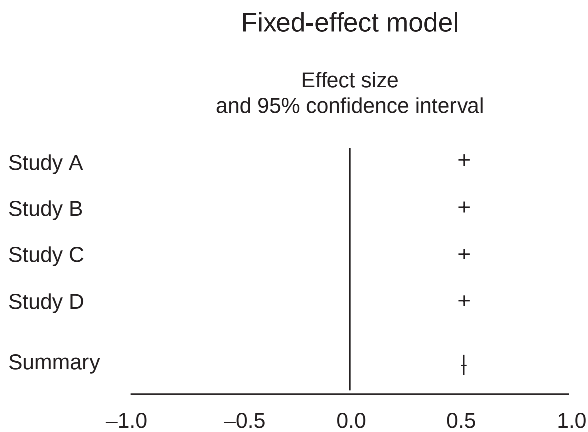

Figure 13.3 Very large studies under fixed-effect model.

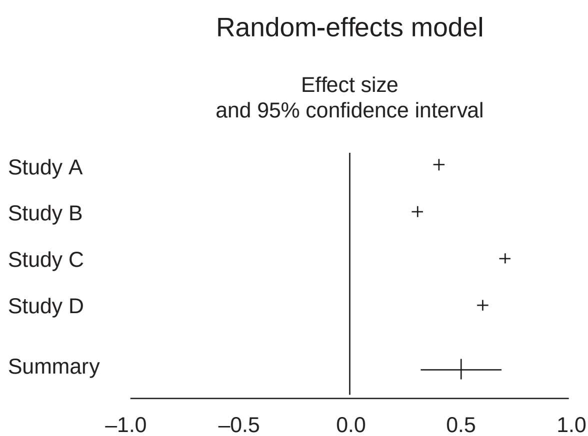

Figure 13.4 Very large studies under random-effects model.

Consider what would happen if we had five studies, and each study had an infinitely large sample size. Under either model the confidence interval for the effect size in each study would have a width approaching zero, since we know the effect size in that study with perfect precision. Under the fixed-effect model the summary effect would also have a confidence interval with a width of zero, since we know the common effect precisely (Figure 13.3). By contrast, under the random-effects model the width of the confidence interval would not approach zero (Figure 13.4). While we know the effect in each study precisely, these effects have been sampled from a universe of possible effect sizes, and provide only an estimate of the mean effect. Just as the error within a study will approach zero only as the sample size approaches infinity, so too the error of these studies as an estimate of the mean effect will approach zero only as the number of studies approaches infinity.

More generally, it is instructive to consider what factors influence the standard error of the summary effect under the two models. The following formulas are based on a meta-analysis of means from k one-group studies, but the conceptual argument applies to all meta-analyses. The within-study variance of each mean depends on the standard deviation (denoted σ) of participants' scores and the sample size of each study (n). For simplicity we assume that all of the studies have the same sample size and the same standard deviation (see Box 13.1 for details).

Under the fixed-effect model the standard error of the summary effect is given by

$$SE_M = \sqrt{\frac{\sigma^2}{k \times n}}. \tag{13.1}$$

It follows that with a large enough sample size the standard error will approach zero, and this is true whether the sample size is concentrated on one or two studies, or dispersed across any number of studies.

Under the random-effects model the standard error of the summary effect is given by

$$SE_M = \sqrt{\frac{\sigma^2}{k \times n} + \frac{\tau^2}{k}}. \tag{13.2}$$

The first term is identical to that for the fixed-effect model and, again, with a large enough sample size, this term will approach zero. By contrast, the second term (which reflects the between-studies variance) will only approach zero as the number of studies approaches infinity. These formulas do not apply exactly in practice, but the conceptual argument does. Namely, increasing the sample size within studies is not sufficient to reduce the standard error beyond a certain point (where that point is determined by τ^2 and k). If there is only a small number of studies, then the standard error could still be substantial even if the total n is in the tens of thousands or higher.

BOX 13.1 FACTORS THAT INFLUENCE THE STANDARD ERROR OF THE SUMMARY EFFECT.

To illustrate the concepts with some simple formulas, let us consider a meta-analysis of studies with the very simplest design, such that each study comprises a single sample of n observations with standard deviation σ. We combine estimates of the mean in a meta-analysis. The variance of each estimate is

$$V_{Y_i} = \frac{\sigma^2}{n}$$

so the (inverse-variance) weight in a fixed-effect meta-analysis is

$$W_i = \frac{1}{\sigma^2/n} = \frac{n}{\sigma^2}$$

and the variance of the summary effect under the fixed-effect model the standard error is given by

$$V_M = \frac{1}{\sum_{i=1}^{k} W_i} = \frac{1}{k \times n/\sigma^2} = \frac{\sigma^2}{k \times n}.$$

Therefore under the fixed-effect model the (true) standard error of the summary mean is given by

$$SE_M = \sqrt{\frac{\sigma^2}{k \times n}}.$$

Under the random-effects model the weight awarded to each study is

$$W^*_i = \frac{1}{\left(\sigma^2/n\right) + \tau^2}$$

and the (true) standard error of the summary mean turns out to be

$$SE_{M^*} = \sqrt{\frac{\sigma^2}{k \times n} + \frac{\tau^2}{k}}.$$

THE NULL HYPOTHESIS

Often, after computing a summary effect, researchers perform a test of the null hypothesis. Under the fixed-effect model the null hypothesis being tested is that there is zero effect in *every study*. Under the random-effects model the null hypothesis being tested is that the *mean effect* is zero. Although some may treat these hypotheses as interchangeable, they are in fact different, and it is imperative to choose the test that is appropriate to the inference a researcher wishes to make.

WHICH MODEL SHOULD WE USE?

The selection of a computational model should be based on our expectation about whether or not the studies share a common effect size and on our goals in performing the analysis.

Fixed effect

It makes sense to use the fixed-effect model if two conditions are met. First, we believe that all the studies included in the analysis are functionally identical. Second, our goal is to compute the common effect size for the identified population, and not to generalize to other populations.

For example, suppose that a pharmaceutical company will use a thousand patients to compare a drug versus placebo. Because the staff can work with only 100 patients at a time, the company will run a series of ten trials with 100 patients in each. The studies are identical in the sense that any variables which can have an impact on the outcome are the same across the ten studies. Specifically, the studies draw patients from a common pool, using the same researchers, dose, measure, and so on (we assume that there is no concern about practice effects for the researchers, nor for the different starting times of the various cohorts). All the studies are expected to share a common effect and so the first condition is met. The goal of the analysis is to see if the drug works in the population from which the patients were drawn (and not to extrapolate to other populations), and so the second condition is met, as well.

In this example the fixed-effect model is a plausible fit for the data and meets the goal of the researchers. It should be clear, however, that this situation is relatively rare. The vast majority of cases will more closely resemble those discussed immediately below.

Random effects

By contrast, when the researcher is accumulating data from a series of studies that had been performed by researchers operating independently, it would be unlikely that all the studies were functionally equivalent. Typically, the subjects or interventions in these studies would have differed in ways that would have impacted on the results, and therefore we should not assume a common effect size. Therefore, in these cases the random-effects model is more easily justified than the fixed-effect model.

Additionally, the goal of this analysis is usually to generalize to a range of scenarios. Therefore, if one did make the argument that all the studies used an identical, narrowly defined population, then it would not be possible to extrapolate from this population to others, and the utility of the analysis would be severely limited.

A caveat

There is one caveat to the above. If the number of studies is very small, then the estimate of the between-studies variance (τ^2) will have poor precision. While the random-effects model is still the appropriate model, we lack the information needed to apply it correctly. In this case the reviewer may choose among several options, each of them problematic.

One option is to report the separate effects and *not* report a summary effect. The hope is that the reader will understand that we cannot draw conclusions about the effect size and its confidence interval. The problem is that some readers will revert to vote counting (see Chapter 28) and possibly reach an erroneous conclusion.

Another option is to perform a fixed-effect analysis. This approach would yield a descriptive analysis of the included studies, but would not allow us to make inferences about a wider population. The problem with this approach is that (a) we do want to make inferences about a wider population and (b) readers will make these inferences even if they are not warranted.

A third option is to take a Bayesian approach, where the estimate of τ^2 is based on data from outside of the current set of studies. This is probably the best option, but the problem is that relatively few researchers have expertise in Bayesian meta-analysis. Additionally, some researchers have a philosophical objection to this approach.

For a more general discussion of this issue see *When does it make sense to perform a meta-analysis* in Chapter 40.

MODEL SHOULD NOT BE BASED ON THE TEST FOR HETEROGENEITY

In the next chapter we will introduce a test of the null hypothesis that the between-studies variance is zero. This test is based on the amount of between-studies variance observed, relative to the amount we would expect if the studies actually shared a common effect size.

Some have adopted the practice of starting with a fixed-effect model and then switching to a random-effects model if the test of homogeneity is statistically significant. This practice should be strongly discouraged because the decision to use the random-effects model should be based on our understanding of whether or not all studies share a common effect size, and not on the outcome of a statistical test (especially since the test for heterogeneity often suffers from low power).

If the study effect sizes are seen as having been sampled from a *distribution* of effect sizes, then the random-effects model, which reflects this idea, is the logical one to use. If the between-studies variance is substantial (and statistically significant) then the fixed-effect model is inappropriate. However, even if the between-studies variance does not meet the criterion for statistical significance (which may be due simply to low power) we should still take account of this variance when assigning weights. If T^2 turns out to be zero, then the random-effects analysis reduces to the fixed-effect analysis, and so there is no cost to using this model.

On the other hand, if one has elected to use the fixed-effect model *a priori* but the test of homogeneity is statistically significant, then it would be important to revisit the assumptions that led to the selection of a fixed-effect model.

CONCLUDING REMARKS

Our discussion of differences between the fixed-model and the random-effects model focused largely on the computation of a summary effect and the confidence intervals for the summary effect. We did not address the implications of the dispersion itself. Under the fixed-effect model we assume that all dispersion in observed effects is due to sampling error, but under the random-effects model we allow that some of that dispersion reflects real differences in effect size across studies. In the chapters that follow we discuss methods to quantify that dispersion and to consider its substantive implications.

Although throughout this book we define a fixed-effect meta-analysis as assuming that every study has a common true effect size, some have argued that the fixed-effect method is valid without making this assumption. The point estimate of the effect in a fixed-effect meta-analysis is simply a weighted average and does not strictly require the assumption that all studies estimate the same thing. For simplicity and clarity we adopt a definition of a fixed-effect meta-analysis that does assume homogeneity of effect.

SUMMARY POINTS

- A fixed-effect meta-analysis estimates a single effect that is assumed to be common to every study, while a random-effects meta-analysis estimates the mean of a distribution of effects.
- Study weights are more balanced under the random-effects model than under the fixed-effect model. Large studies are assigned less relative weight and small studies are assigned more relative weight as compared with the fixed-effect model.
- The standard error of the summary effect and (it follows) the confidence intervals for the summary effect are wider under the random-effects model than under the fixed-effect model.

- The selection of a model must be based solely on the question of which model fits the distribution of effect sizes, and takes account of the relevant source(s) of error. When studies are gathered from the published literature, the random-effects model is generally a more plausible match.
- The strategy of starting with a fixed-effect model and then moving to a random-effects model if the test for heterogeneity is significant is a mistake, and should be strongly discouraged.